

Linking Fidelity of Implementation to Outcomes in Real-World Settings

by *Meghan McCormick*

AUGUST 2019

This post is one in a series highlighting MDRC's methodological work. Contributors discuss the refinement and practical use of research methods being employed across our organization.

Part I of this two-part post discussed MDRC's work with practitioners from the Boston Public Schools (BPS) Department of Early Childhood to construct valid and reliable measures of implementation fidelity to an early childhood curriculum. This Part II examines how those data can reveal associations between levels of fidelity and gains in children's academic skills.

Researchers testing the effects of interventions in education and social policy are often interested in the fidelity of implementation — that is, whether a program is being implemented as designed. If fidelity to the program model is high, then the evaluation is a “true test” of the intended intervention. If fidelity to the program model is low, then if the study does not detect an impact of a meaningful size (assuming the study is well designed and has sufficient statistical power) it may be because the model was not implemented as intended.

Although some studies include substantial resources to promote high levels of fidelity, others are conducted under real-world conditions by entities such as schools and community-based organizations. As a result, the fidelity of implementation can vary widely across settings. This natural variation in implementation across settings can be used to test whether impacts are greater at sites with stronger fidelity. Such findings can be used to understand how important fidelity is to improving outcomes of interest.

An example of such a situation emerged in MDRC's partnership with the BPS Department of Early Childhood, the Harvard Graduate School of Education, and the University of Michigan. This partnership is examining the roll-out of a new curriculum called *Focus on Early Learning*, developed by BPS to improve instruction from preschool to second grade. As discussed in Part I, the research team worked with the school district to construct a classroom observation tool measuring fidelity to the curriculum. A primary goal of this work was to inform program-improvement efforts by identifying the practices whose successful implementation was most strongly linked to improvements in students' outcomes, as well as the schools and classrooms that appeared to struggle the most to implement those practices. The school district and research team were particularly interested in learning more about students' gains over time, and about whether students enrolled in preschool classrooms with higher fidelity to the curriculum



mdrc
BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

REFLECTIONS
ON
METHODOLOGY

NEW YORK
200 Vesey Street, 23rd Floor
New York, NY 10281
Tel: 212 532 3200

OAKLAND
475 14th Street, Suite 750
Oakland, CA 94612
Tel: 510 663 6372

WASHINGTON, DC
1990 M Street, NW, Suite 340
Washington, DC 20036

LOS ANGELES
11965 Venice Boulevard, Suite 402
Los Angeles, CA 90066

www.mdrc.org

demonstrated faster growth in math and language skills than students in classrooms with lower fidelity.

One challenge in this work was to create a small number of measures that could capture indicators of fidelity observed during a live classroom observation that lasted about two hours. The team grouped observed indicators into four overarching composite measures, or constructs:

- 1 **Effectively extending and building on students' current language and math skills** by integrating advanced content into instruction. For example, this construct included observations related to whether “explanations and demonstrations that build conceptual knowledge are the teacher’s dominant instructional strategies.”
- 2 **Scaffolding learning** (meaning that teachers recognize children’s current skill levels and effectively support them in moving to the next level) and **differentiating instruction** (so that it is tailored to the skills, knowledge, and interests of individual students). For example, this construct included observations related to whether the “teacher adapted the task or discussion according to children’s abilities and development by purposefully presenting the content in different ways, varying materials, or providing children with flexibility in how they complete the activity(ies).”
- 3 **Summarizing content and making connections** both across content areas and in students’ lives. For example, this construct included observations related to whether the “teacher verbally summarizes/reflects on the lesson before transitioning to the next activity.”
- 4 **Use of rich vocabulary in instruction** with specific, developmentally appropriate, advanced vocabulary words embedded into instruction, as defined by the curriculum. For example, this construct included observations related to whether the “teacher is intentional in which vocabulary words are used and how they are defined.”

As discussed in Part I, this effort demonstrated the importance of capturing similar types of information across observations, which can be difficult when collecting data in

real-world settings because of variation in teachers’ schedules, the curricular components that can be seen on a particular day, and time constraints in gathering observational data. Using the information collected, the team used a series of analyses to show that the items in these constructs were related to one another, that the constructs were not redundant with one another, and that the constructs were moderately correlated with other measures of classroom quality.

After completing this measurement work, the team used multilevel models with students nested within classrooms and schools to examine how growth in math and language scores in the spring of the preschool year were associated with these four fidelity constructs, controlling for a robust set of demographic characteristics. These baseline control variables helped to account for the potential selection of students into higher- versus lower-fidelity settings. The team found evidence that Hispanic and dual language learner students who experienced higher fidelity to the *Focus on Early Learning* curriculum in preschool demonstrated faster growth in math skills than their similar peers who experienced lower fidelity to the curriculum. (The results of this study cannot be interpreted causally and only reveal associations between implementation fidelity and children’s outcomes.)

Hispanic students make up 36 percent of the BPS preschool population; dual language learner students make up 52 percent. These findings could therefore help BPS understand one way it might be possible to improve outcomes for these two important groups. The district has decided to try to improve fidelity in schools with weak implementation, focusing particularly on schools serving high percentages of Hispanic and dual language learner students.

The work described in this post continues, and the team is currently trying to replicate these results with kindergarten and first-grade data. The findings suggest that fidelity data can be used outside of research demonstrations to aid in program improvement and link program features to gains in outcomes. Such efforts may eventually lead to the type of experimental evaluation that can inform policy and practice more rigorously.